

# 統計計算與模擬

政治大學統計系余清祥

2026年3月11日

第二單元：隨機變數模擬

<http://csyue.nccu.edu.tw>



# 以下分配的亂數如何產生？

The 10 examples of Discrete Random Variables are the following;

1. The number of outcomes of tossing a fair coin.
2. The number of students inside the classroom.
3. The number of honors during the school year.
4. The number of covid cases on a daily basis.
5. The number of patients in a ward.
6. The number of vaccine dosages.
7. The number of eggs sold in a day.
8. The number of recoveries from Novel Corona Virus 19 in a week.
9. The number of equations used to solve a problem.
10. The number of items during the examination.

The 10 examples of Continuous Random Variables are the following;

1. The distance from your school to your home.
2. The minimum salary of an employee.
3. The height requirement to become a flight attendant.
4. The minimum weight before obesity.
5. The average grades you during a semester.
6. The amount you invested for the future.
7. The temperature during the wet season.
8. The minimum temperature to store the vaccines.
9. The amount of water that a box can contain.
10. The amount of air pressure in a tank.

# 常見的隨機變數產生方式

(Random Numbers from certain Distribution)

均勻分配以外的隨機變數通常藉由下列方法，透過均勻亂數產生。

- Inverse Transform Method
- Composition Method
- Rejection (and Acceptance) Method
- Alias Method
- Table Method

註：二維以上變數的模擬也類似，但需要矩陣的輔助，將在下一單元介紹。

# Normal Distribution

- 與均勻分配類似，常態分配是最常用到的分配，常見的常態分配亂數產生法：

(Random numbers from normal distribution is one of the popular choices for simulation.)

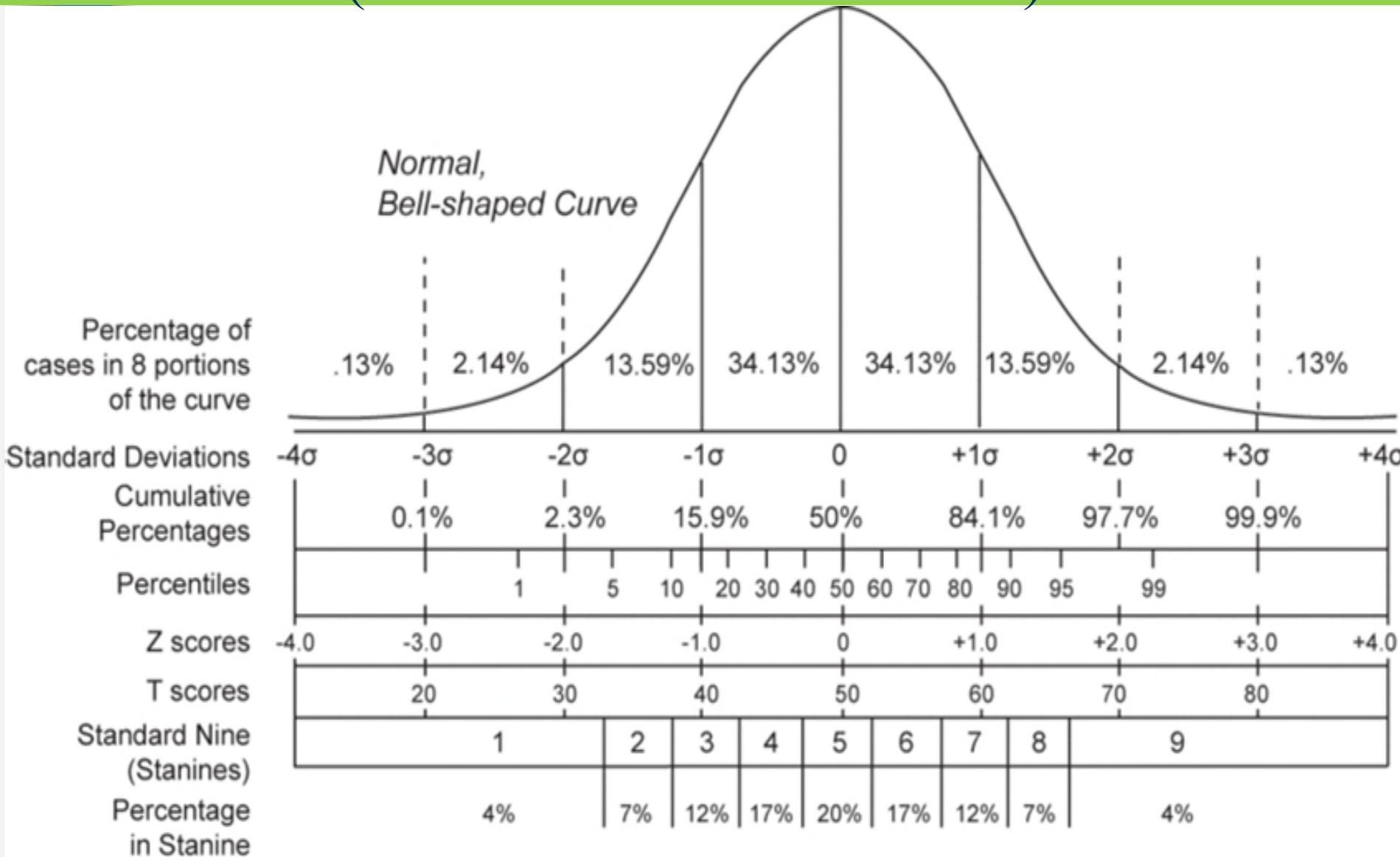
→  $\sum_{i=1}^{12} U_i - 6$  (僅為近似，only approximation)

→ Box-Muller

→ Polar

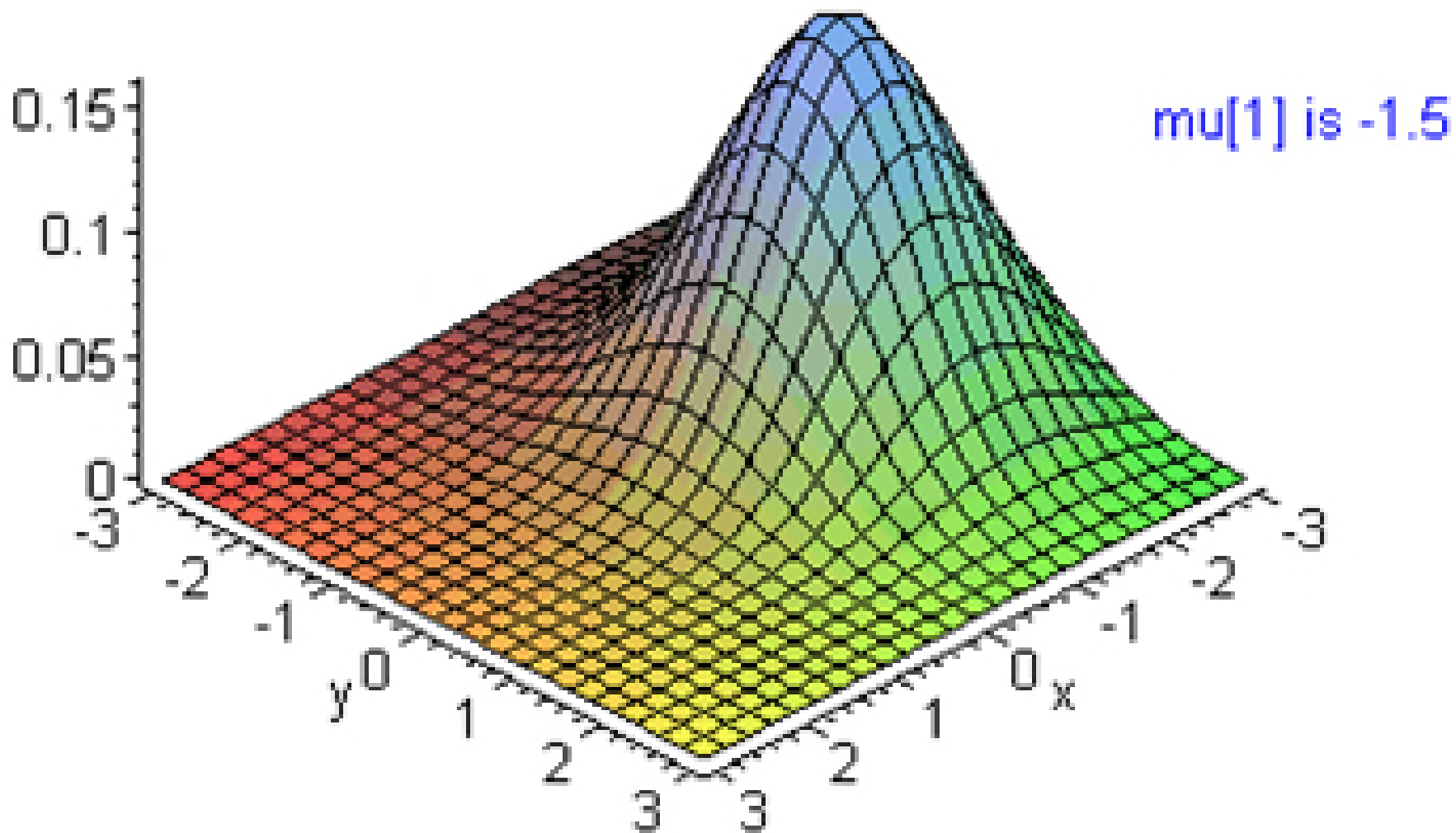
→ Ratio-of-uniforms

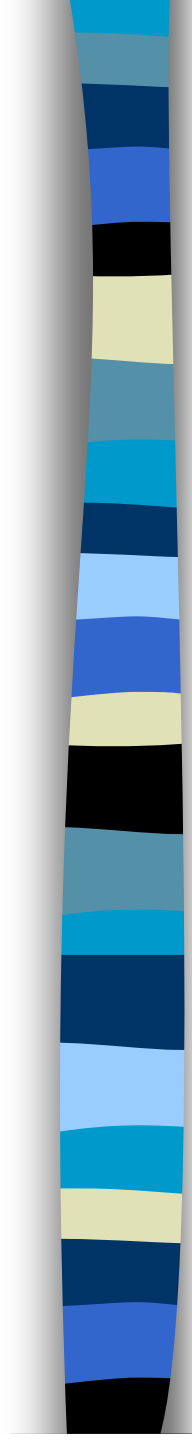
# 常態分配的重要特性 (About normal distribution)



# 二維與多維常態分配

(Bivariate and Multivariate Normal Distribution)





- $Y = \sum_{i=1}^{12} U_i - 6$  (近似法)

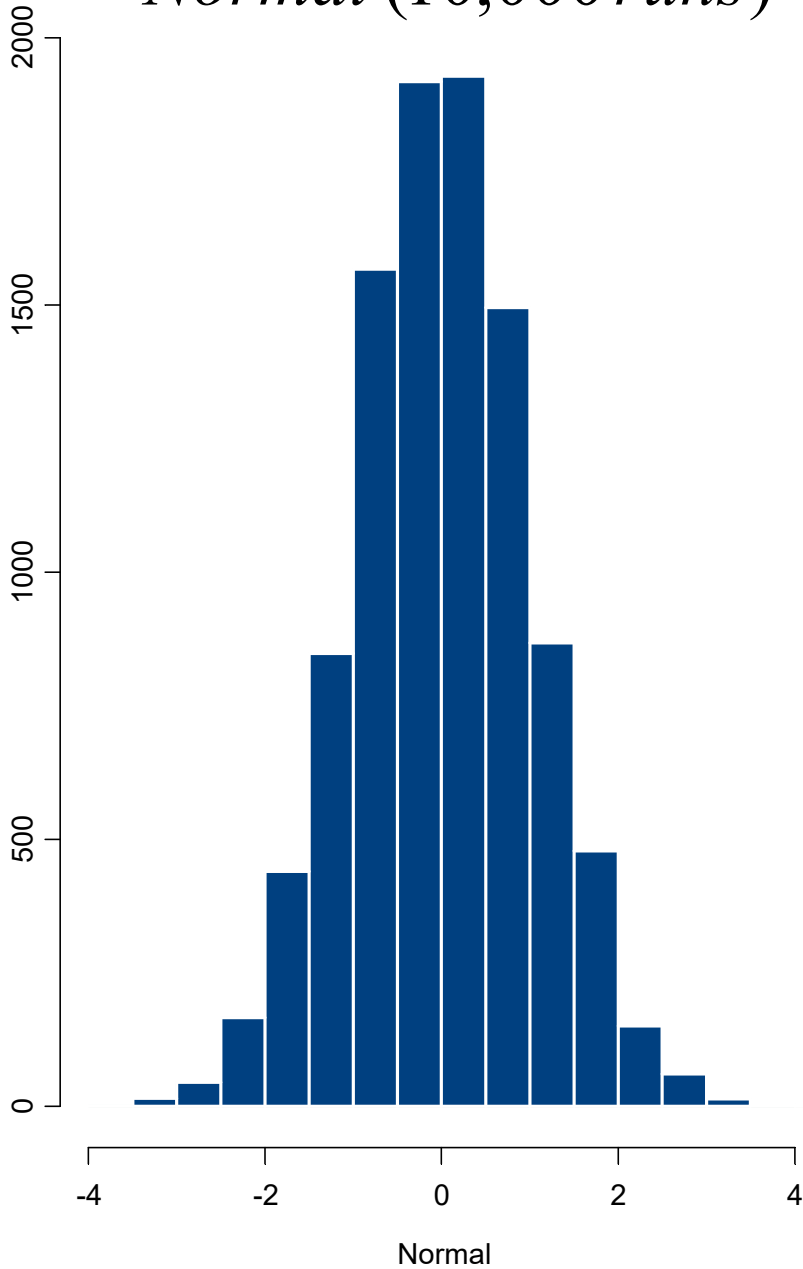
→ 因為  $Y$  的期望值與變異數等於

$$E(Y) = E\left(\sum_{i=1}^{12} U_i\right) - 6 = 6 - 6 = 0$$

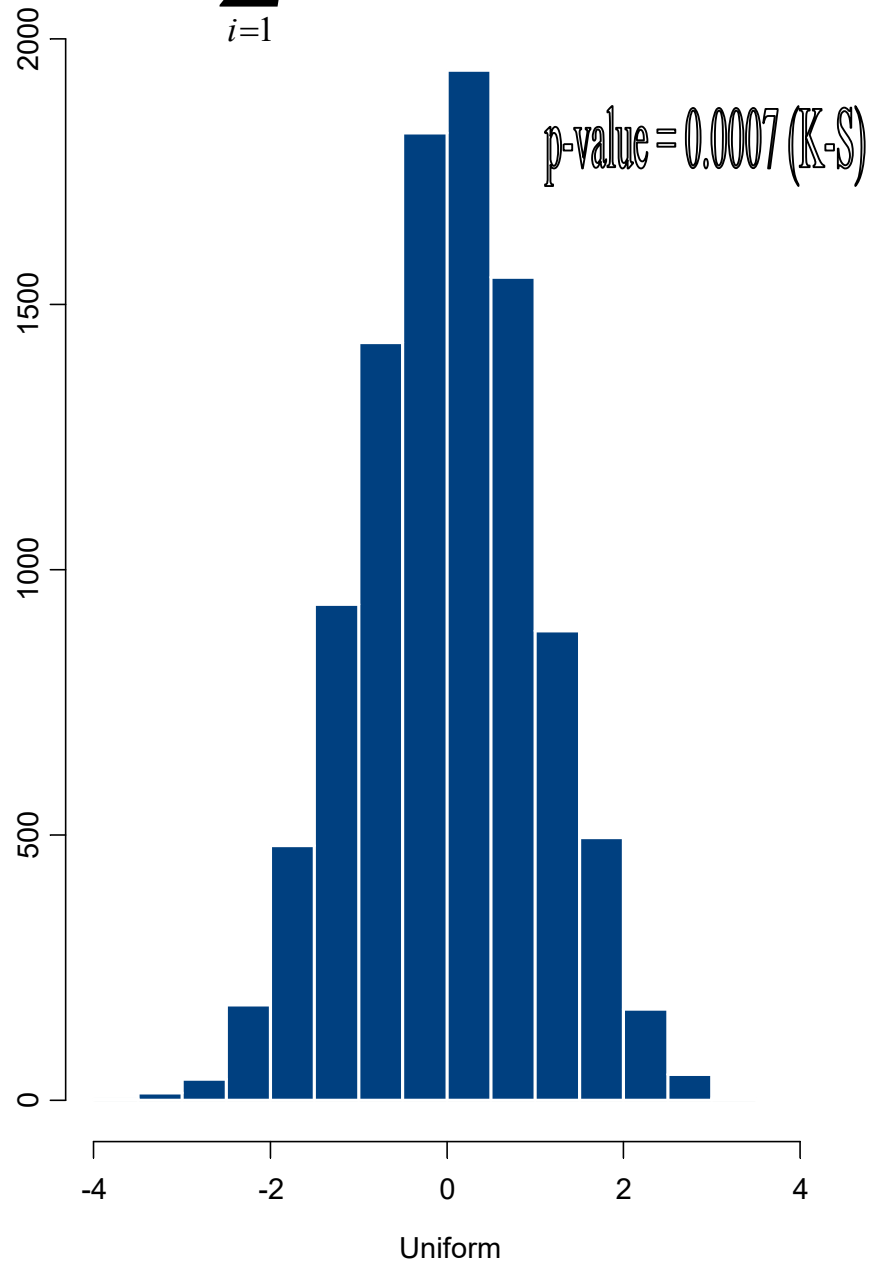
$$Var(Y) = 12 \cdot Var(U_i) = 12 \cdot \frac{1}{12} = 1$$

與標準常態分配相同。

*Normal (10,000 runs)*



$$\sum_{i=1}^{12} U_i - 6 \text{ (10,000 runs)}$$





- Box and Muller (1958)

→ The best known “exact” method for the normal distribution.

- Algorithm

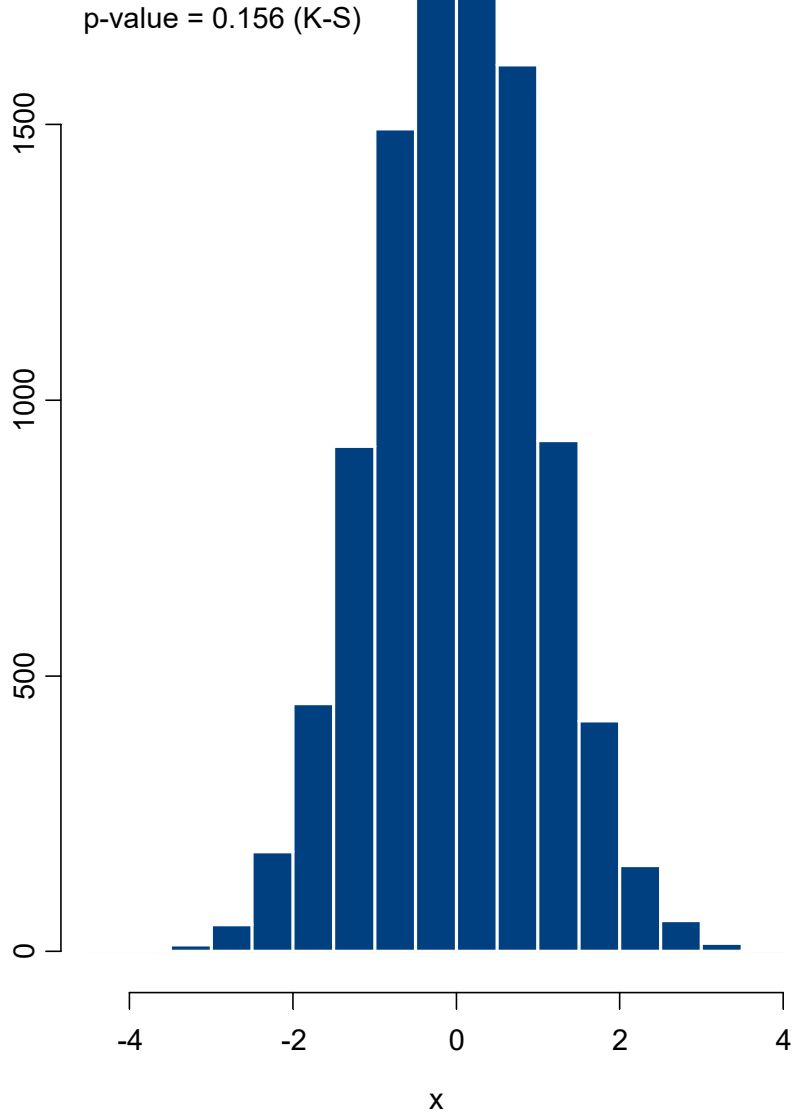
1. Generate  $U_1, U_2 \sim U(0, 1)$

2. Let  $\theta = 2\pi U_1$

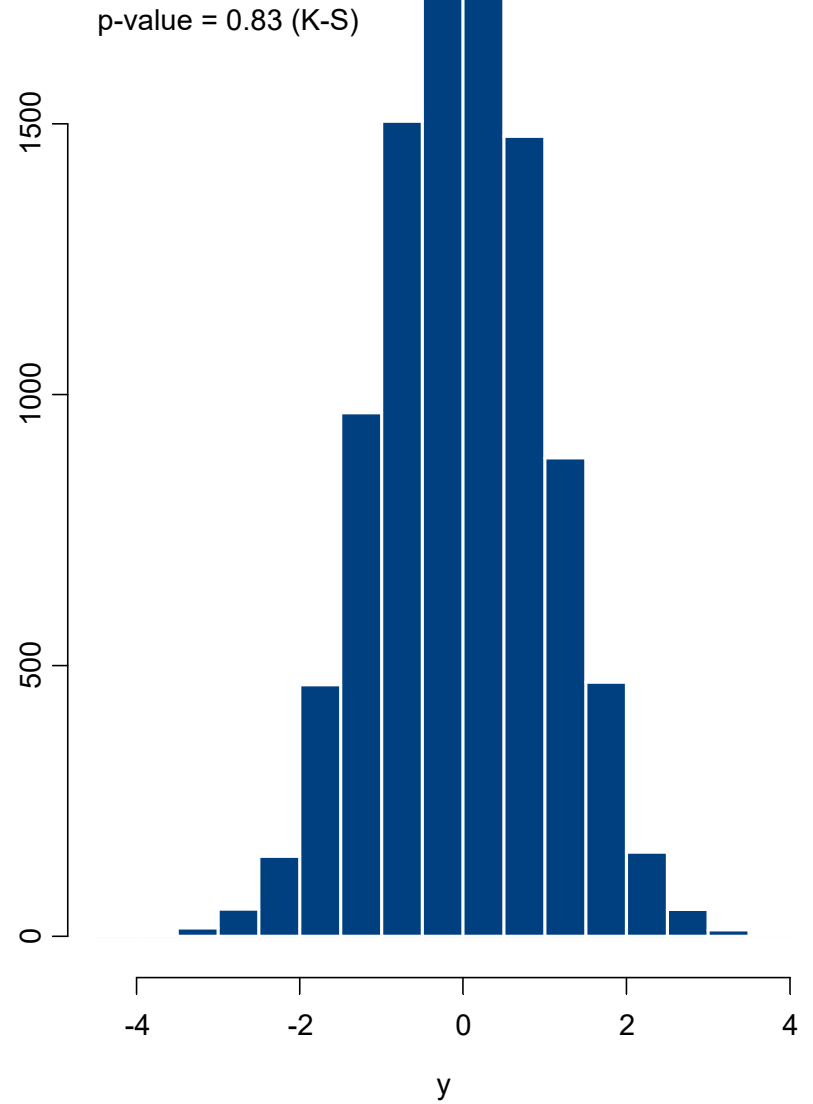
$$E = -\log U_2 \quad \& \quad R = \sqrt{2E}$$

3. Then  $X = R \cos \theta$  and  $Y = R \sin \theta$  are independent standard normal variables.

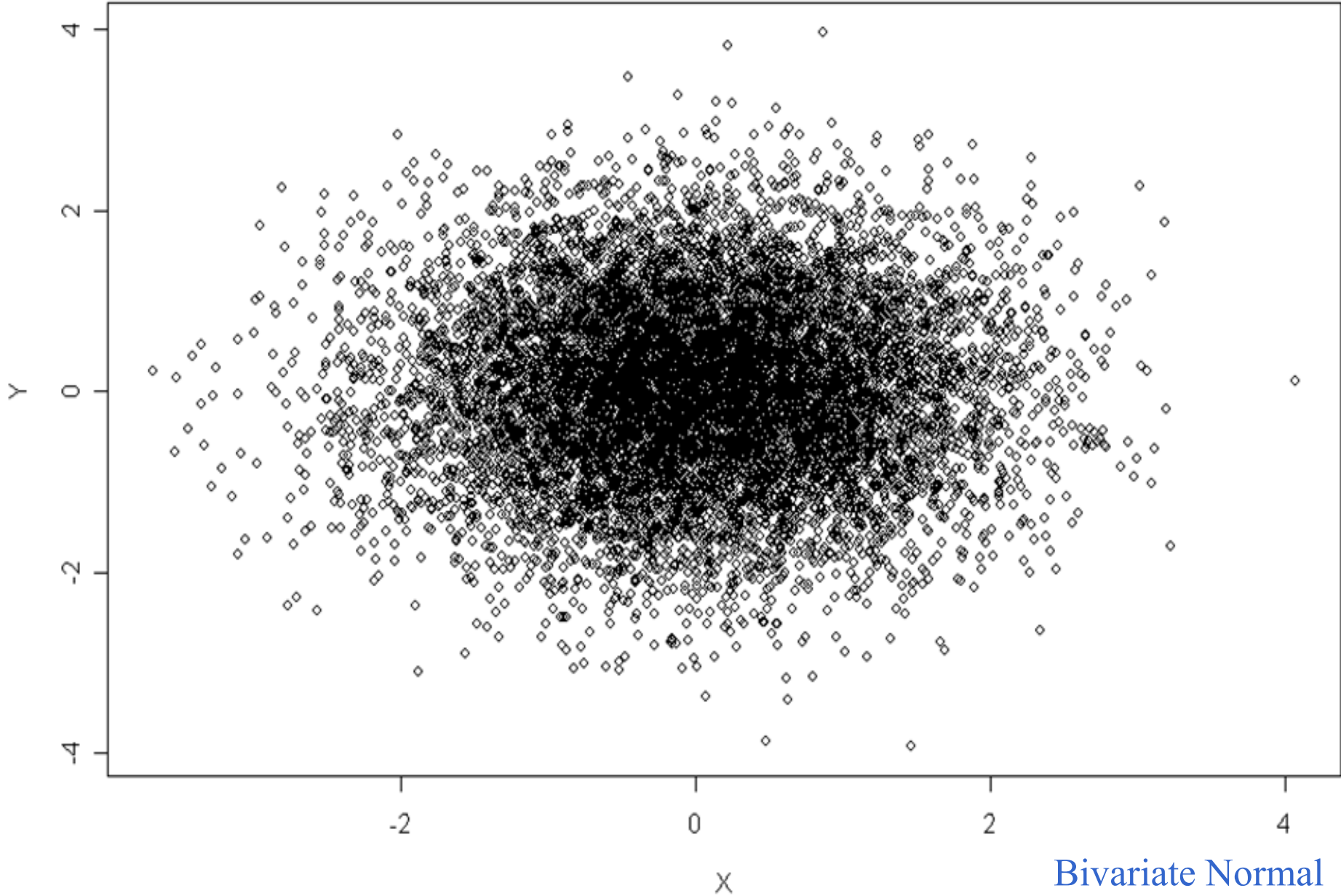
Box-Muller (10,000 runs)



correlation(X,Y)=-0.0039



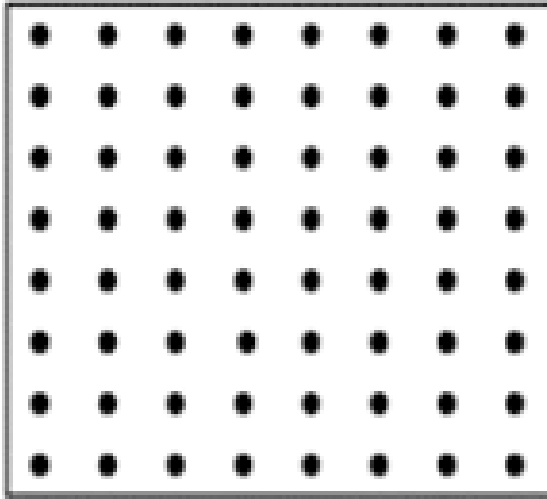
# Scatter Plot of Box-Muller



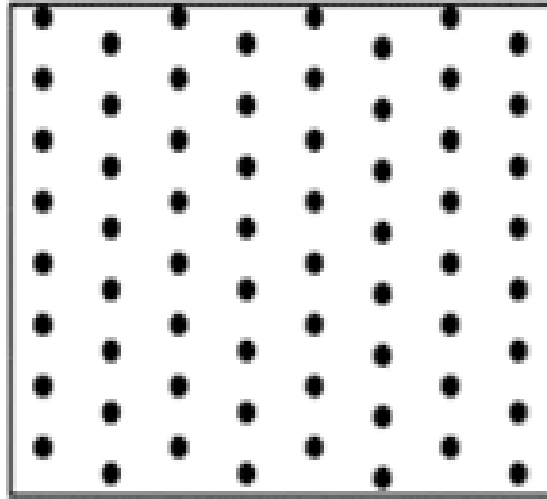
Bivariate Normal

# 二維觀察值的可能特性

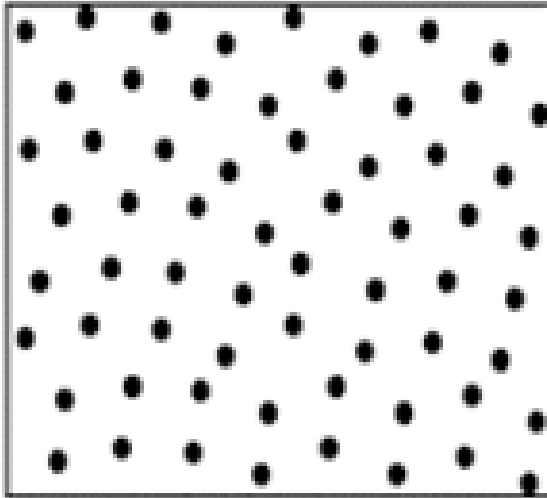
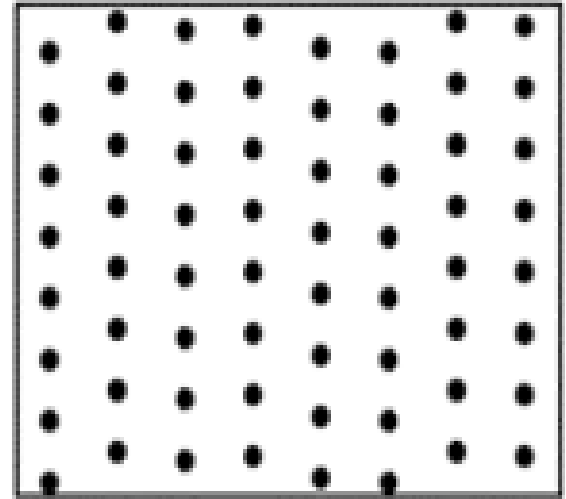
REGULAR



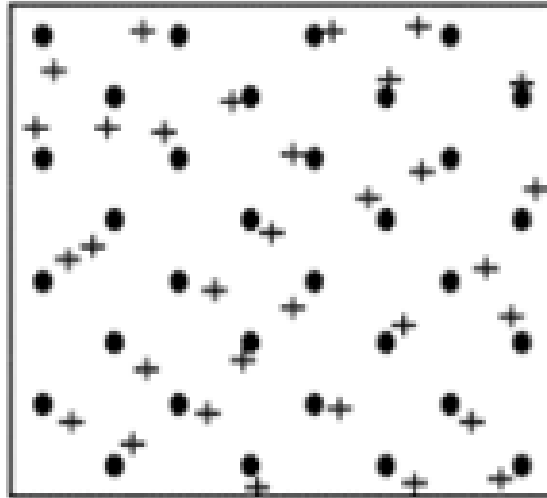
STAGGERED START



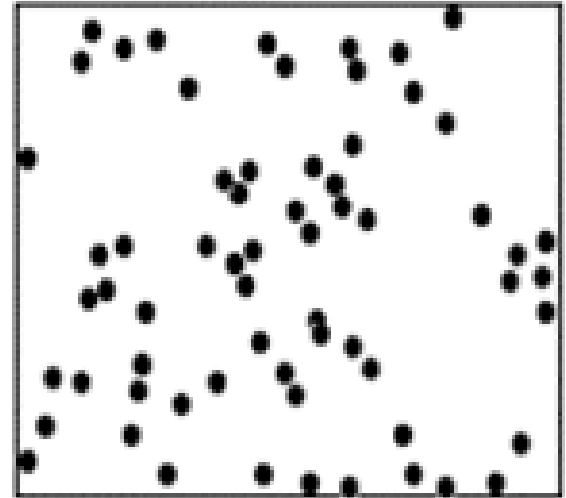
RANDOM START



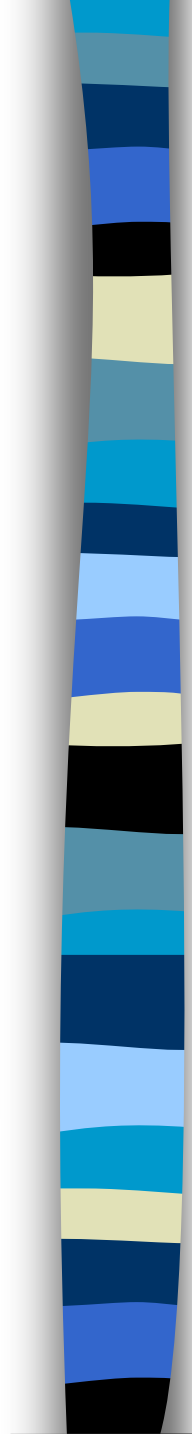
SYSTEMATIC UNALIGNED



RANDOM CLUSTER



SIMPLE RANDOM

- 
- Note: Using Box-Muller method with congruential generators must be careful. It is found by several researchers that

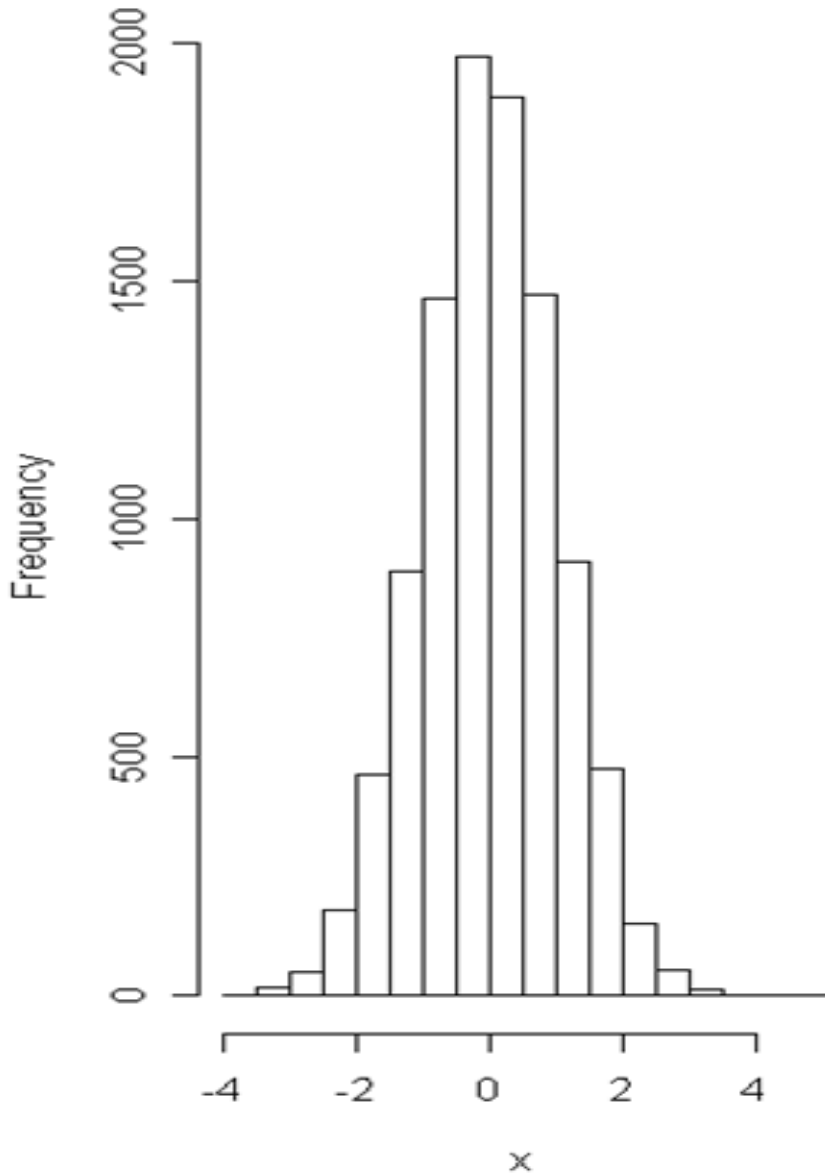
$$a \text{ (乘數)} = 131$$

$$c \text{ (增量)} = 0$$

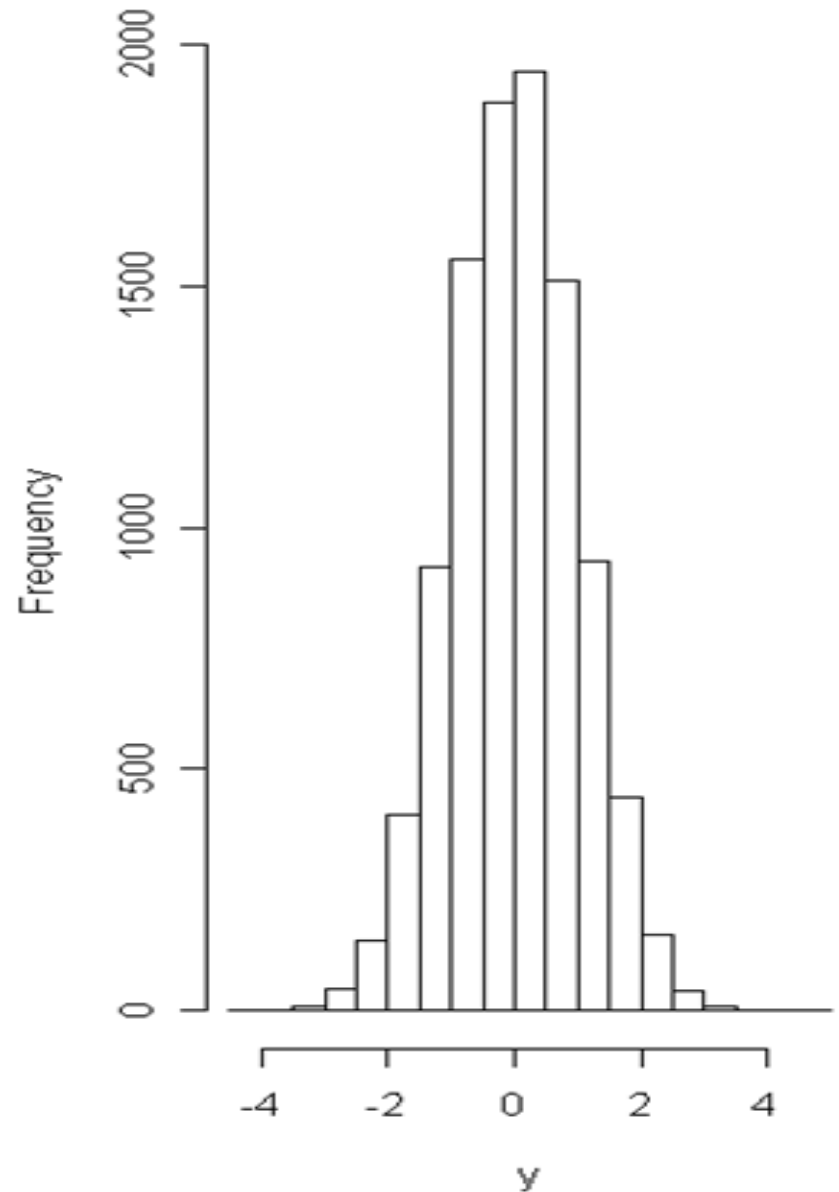
$$m \text{ (除數)} = 2^{35}$$

would have  $X \in (-3.3, 3.6)$ . See Neave (1973) and Ripley (1987, p.55) for further information.

**a=131, c=0, m=2<sup>35</sup>**



**Box-Muller**



I use 123,456 and 3,456 as seeds but did not find problems.



- Polar Method (recommended!)

→ Rejection method for generating two independent normal variables.

- Algorithm

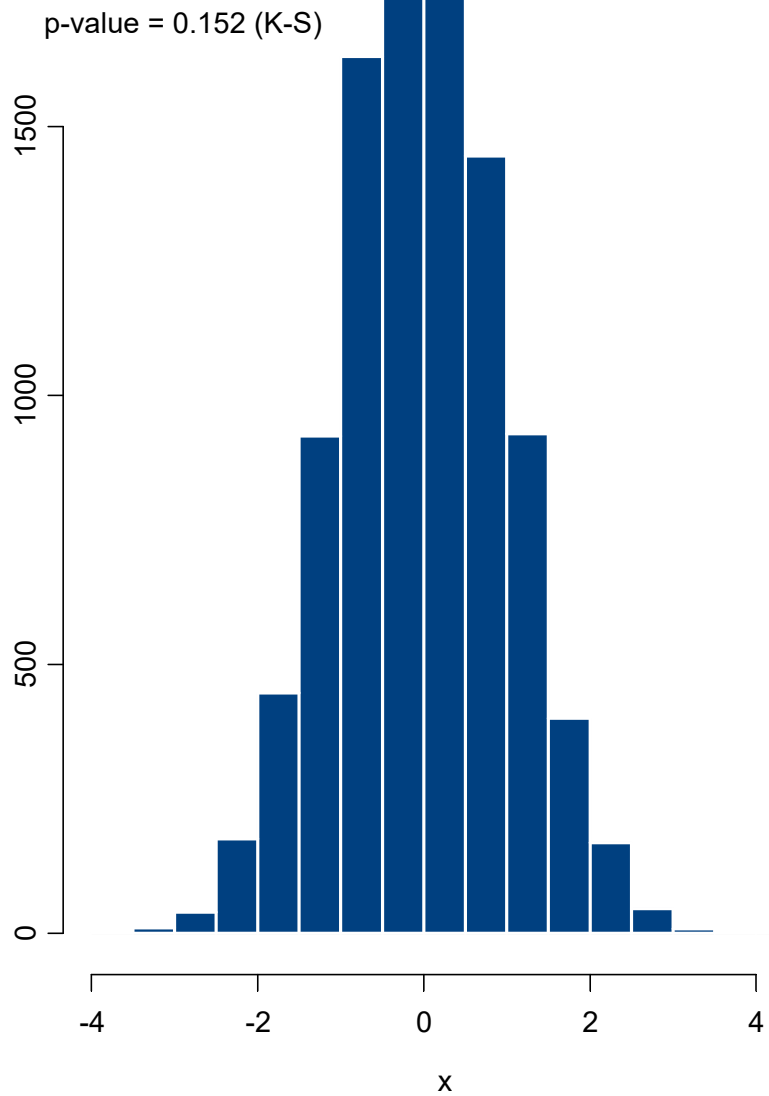
1. Generate  $V_1, V_2 \sim U(-1, 1)$

retain if  $W = V_1^2 + V_2^2 < 1$

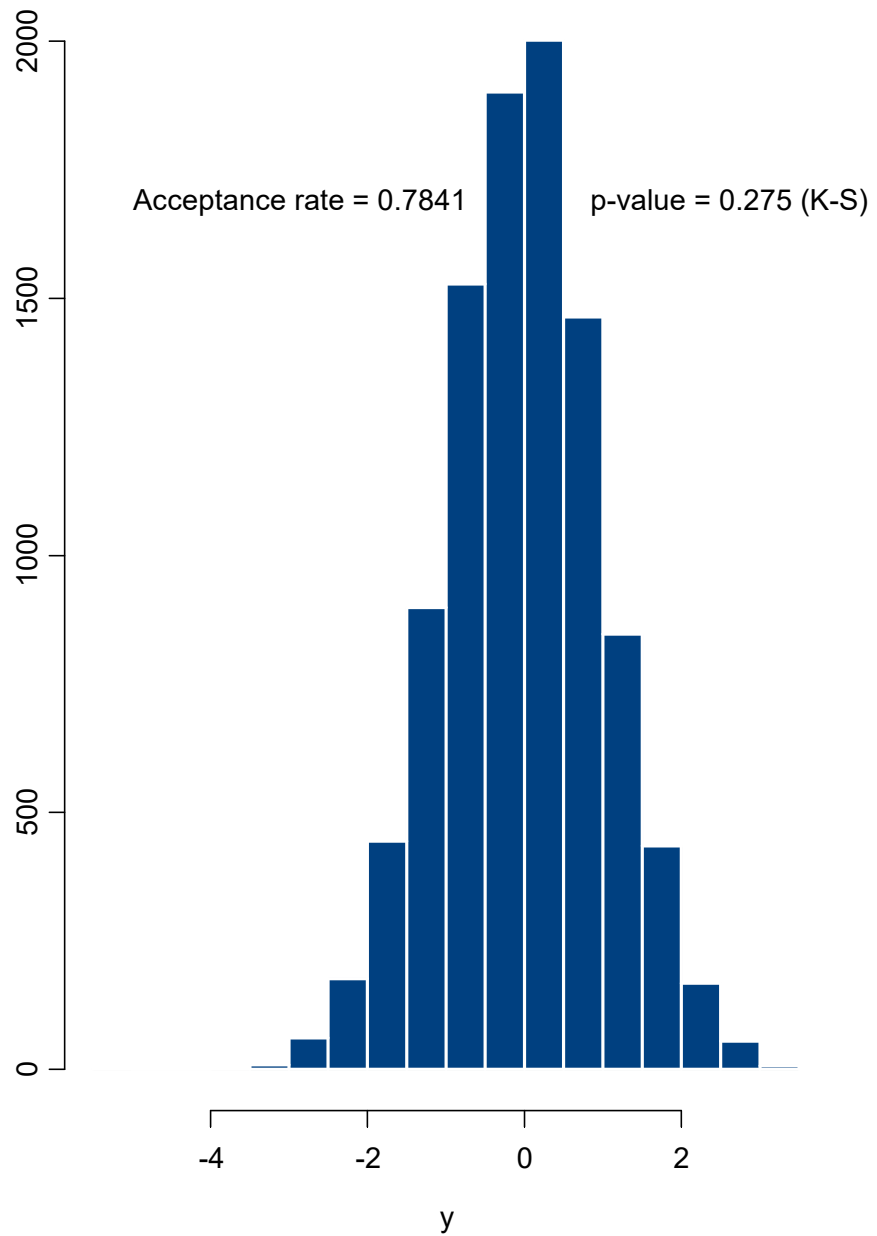
2. Let  $C = \sqrt{-2W^{-1} \log W}$

3. Then  $X = CV_1$  and  $Y = CV_2$  are independent standard normal variables.

Polar method (10,000 runs)



Correlation(X,Y)=0.0087





- Ratio-of-uniforms (recommended!)

→ Similar to Polar method, Ratio-of-uniforms is a rejection method.

- Algorithm

1. Generate  $U_1, U_2 \sim U(0, 1)$ ,

*and let*  $V = \sqrt{2/e} (2U_2 - 1)$

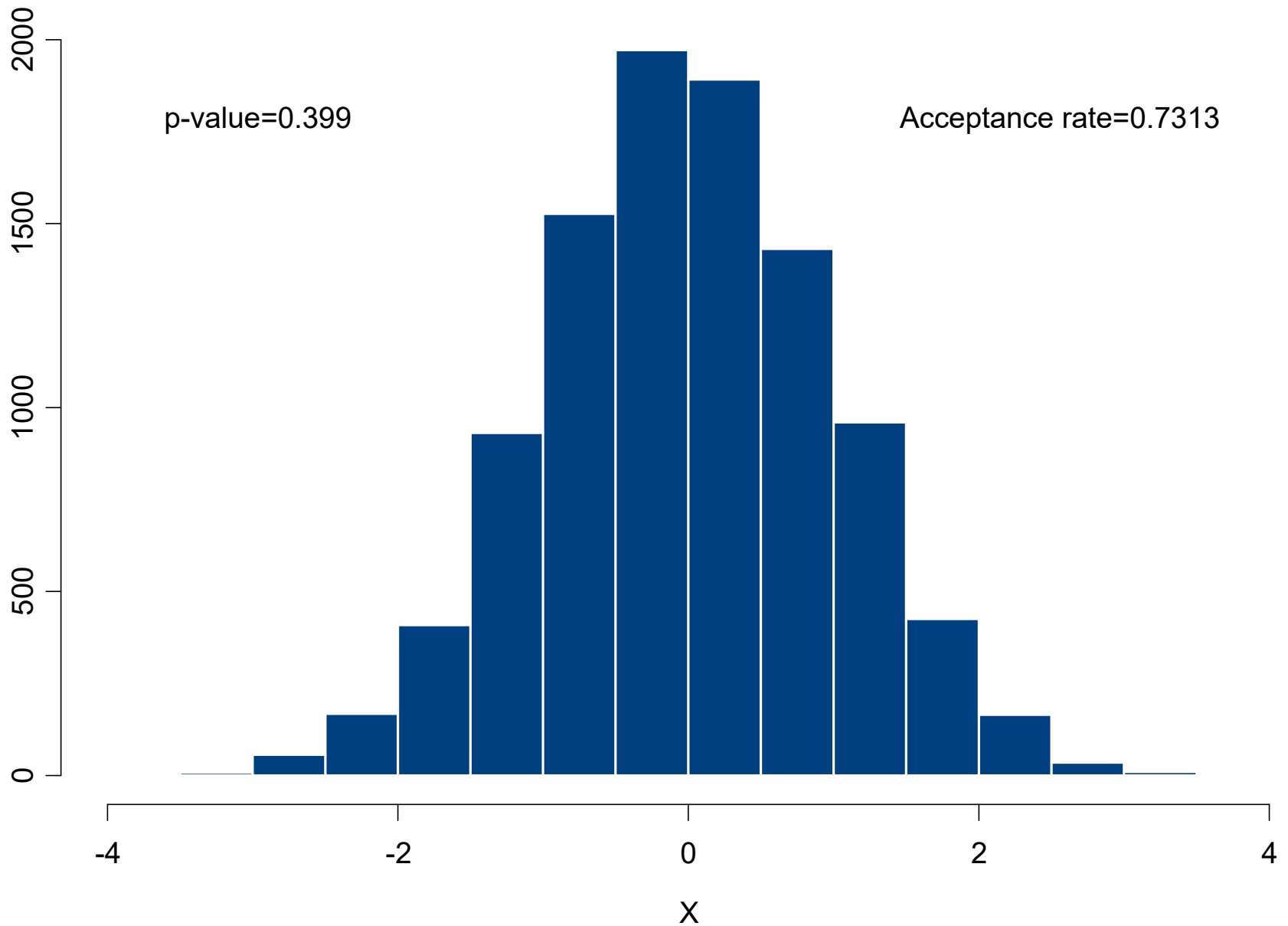
2. Let  $X = V / U_1$ ,  $Z = X^2/4$

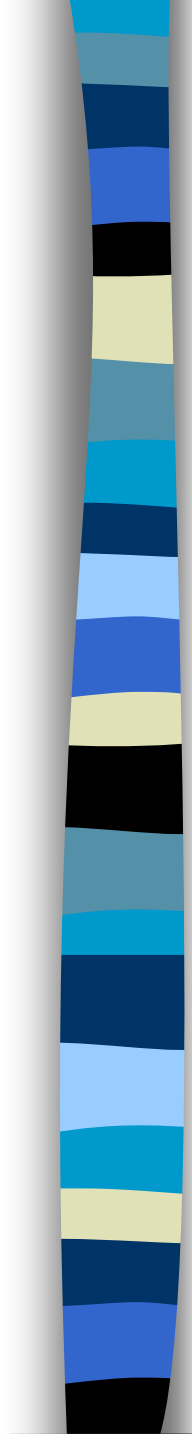
3. Retain if  $Z < 1 - U_1$  → Suggest deleting this step!

4. Retain if  $Z \leq 0.259 / U_1 + 0.35$  &  $Z \leq -\log U_1$

5. Then  $X$  is normally distributed.

# Ratio-of-uniforms (10,000 runs)





■ 問題：隨機變數有不同產生方法時，你/妳會選擇哪一種？(Choices of Generation)

(有哪些判斷標準可供參考？)

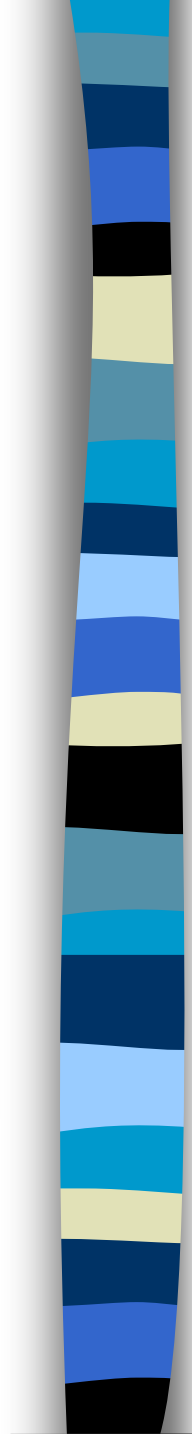
→ 產生效率(e.g.成功接受率)？

→ 方法的複雜/方便程度？

→ 方法是否有瑕疵？(例如：與線性同餘法配合時是否有瑕疵？)

→ 取得的方便/廣為接受的方法？

→ 如何綜合判斷及考量？



■ 是否有需要開發新的隨機變數產生方法？

→ 除非有特別需求(例如：學術研究、軟體配合等)，或是目標的隨機變數較為特殊，建議採用已經廣為大家認可的方法，可省卻驗證方法是否有效，也較不易遭受質疑。

註：如果其他軟體可以產生需要的變數，建議可由該軟體輸出亂數、再輸出，或者參考該軟體產生亂數的語法。(註：可能語法不同產生問題！)



- Ratio of Uniform (Cauchy Distribution)

→ Ratio-of-uniforms can also be used to create random variables from Cauchy Distribution.

- Algorithm

1. Generate  $U_1, U_2 \sim U(0,1)$ , and  $V = 2U_2 - 1$   
retain if  $U_1^2 + V^2 < 1$ .

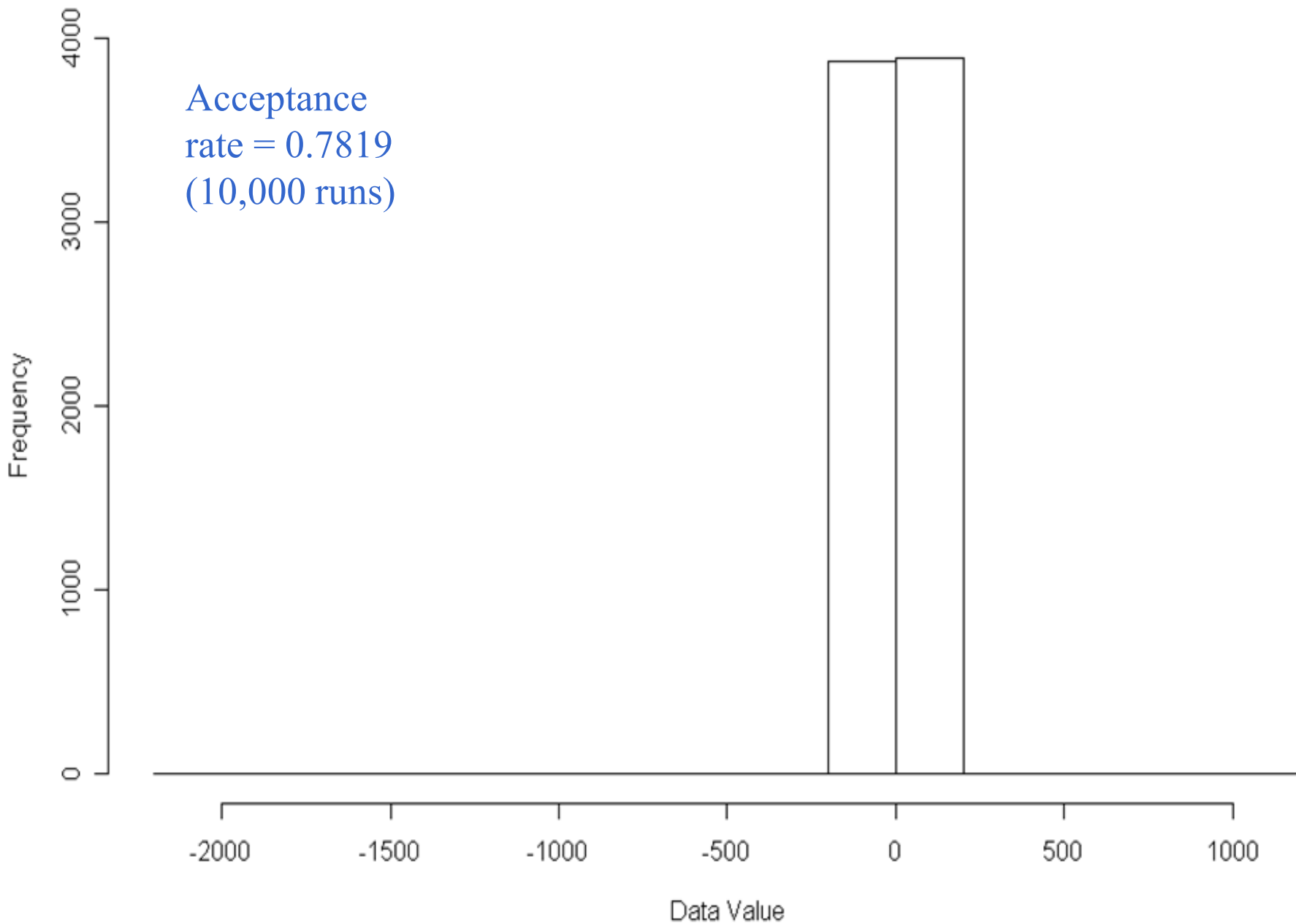
2. Then  $X = V / U_1 \sim \text{Cauchy}(0,1)$ , i.e.

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

e.g. 10,000 runs → p-value=0.9559

(Acceptance rate = 0.7819)

# Histogram of Ratio-of-Uniform (Cauchy)



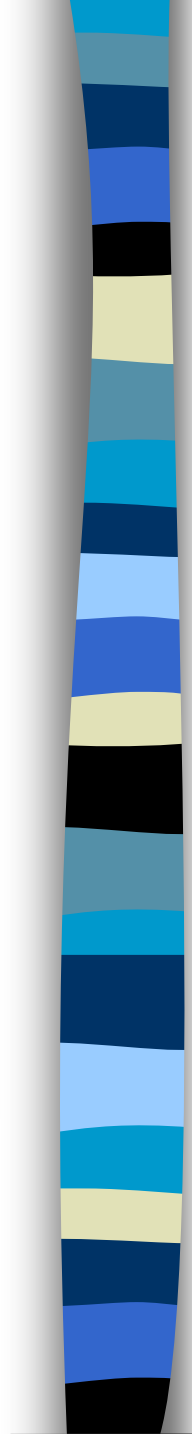


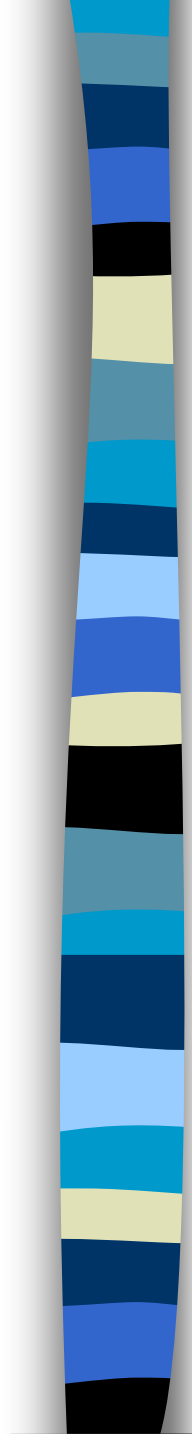
# Inverse Transformation Method

Generates a random number from a probability density function by solving the probability.

Density function's variable in terms of randomly generated numbers. This is achieved as follows:

- We solve the inverse of the integral of our probability density function at an arbitrary point  $F(a)$ , in terms of a random number  $r$ .
- We generate a unique random variable  $a$ , as follows:  $a = F^{-1}(r)$ . The foundation of this method is  $F(X) \sim U(0, 1)$  for all  $X$ .

- 
- Note: Usually, we only use Inversion to create simpler r.v.'s, such as exponential distribution (i.e.  $Exp(\lambda) = -\lambda \log(U(0,1))$ ). Although Inversion is a universal method, it may be too slow (unless subprograms to calculate  $F^{-1}$  are available).
  - In other words, although theoretically it is possible to create any r.v.'s, usually there are simpler methods.



■ Example 1.  $F(x) = x^2$ ,  $0 < x < 1 \rightarrow X = U^{1/2}$ .

■ Example 2. Let  $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} F(x)$   
and we want to generate the minimum  
and maximum of  $X$ 's.

$\rightarrow$  Since  $Y_n = \max(X_1, \dots, X_n)$  &  $Y_1 = \min(X_1, \dots, X_n)$   
which means  $F_{y_n}(y) = [F(y)]^n$  &  $F_{y_1}(y) = 1 - [1 - F(y)]^n$ ,  
we can generate these two r.v.'s by

$$Y_n = F^{-1}(U^{1/n}) \text{ \& } Y_1 = F^{-1}(1 - U^{1/n}).$$

Or,  $Y_n = U^{1/n}$  &  $Y_1 = 1 - U^{1/n}$  if  $X \sim U(0,1)$ .

- 
- Example 3. Generate  $X \sim \text{Poisson}(\lambda)$ , i.e.

$$F(i-1) = P(X \leq i-1) < u \leq P(X \leq i) = F(i),$$

where  $i \in \{0, 1, 2, \dots\}$ .

- Algorithm:

1. Generate  $U_1 \sim U(0, 1)$  and let  $i = 0$ .

2. If  $U_1 \geq F(i)$ , let  $i = i + 1$ ;

Otherwise  $X = i$ .

Note: This method can be used to generate any discrete distributions.



## Notes:

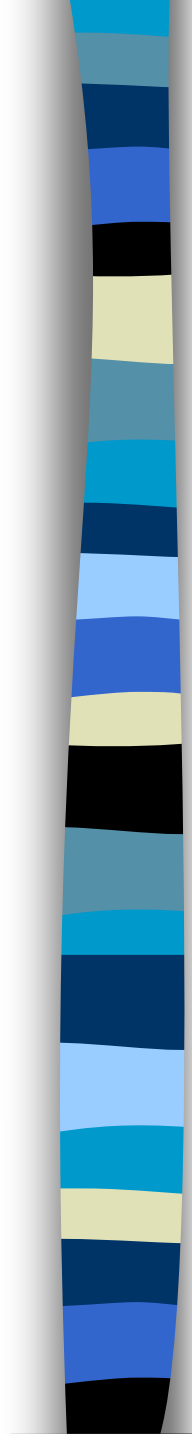
- (1) If total number of classes  $\geq 30$ , we start from the middle (or mode).
- (2) The expected number of trials is  $E(X)+1$ .
- (3) We can use “*Indexed Search*” to increase the efficiency:

→ Fix  $m$ , let  $q_j = \min \{i \mid F(i) \geq \frac{j}{m}\}$ ,  $j = 0, \dots, m-1$

Step 1. Generate  $U \sim U(0, 1)$ , let  $k = [mU]$  &  $i = q_k$

Step 2. If  $U \geq F(i)$ , let  $i = i + 1$ ;

Otherwise  $X = i$ .



■ Example 4.  $X \sim \text{Poisson}(10)$

- (1) Usual method: 11 comparisons on average
- (2) Mode: Reduced to about 3.54 comparisons
- (3) Indexed search: we choose  $m = 5$ , i.e.

$$q_0 = 0, q_1 = 7, q_2 = 9, q_3 = 11, q_4 = 13.$$

Under 10,000 simulation runs (S-Plus), I found that one check on the table plus 2.346 comparisons  $\rightarrow$  3.346

(Comparing to 3.3 in Ripley's book)



# Composition Method

- We can generate complex distribution from simpler distributions, i.e.

$$F(x) = \sum_{i=1}^m \alpha_i F_i(x),$$

where  $F_i(x)$  are d.f. of other variables.

→ Or, from a conditional distribution,

$$f(x) = \sum_i p_i g(x | y = i)$$

$$f(x) = \int g(x | y) dF_Y(y).$$

- 
- Example 1. To simulate  $x$ , where

$$P(x = i) = \begin{cases} 0.05, & i = 1, 2, 3, 4, 5 \\ 0.15, & i = 6, 7, 8, 9, 10 \end{cases}$$

We independently generate  $x_1 \in U\{1, 2, \dots, 10\}$  and  $x_2 \in U\{6, 7, \dots, 10\}$ . Let  $X = X_1$  or  $X_2$  with probability 0.5.

- Example 2.  $X \sim B(5, 0.2)$

→ Sampling on each digit:

$$P(X = i) = \begin{cases} 0.3277, & i = 0 \\ 0.4096, & i = 1 \\ 0.2048, & i = 2 \\ 0.0512, & i = 3 \\ 0.0064, & i = 4 \\ 0.0003, & i = 5 \end{cases}$$



- Table method: (Composition)

Example 1.  $X \sim B(3, 1/3)$

	位置	$10^{-1}$	$10^{-2}$	$10^{-3}$
$P(X=0)=0.296$		2	9	6
$P(X=1)=0.445$		4	4	5
$P(X=2)=0.222$		2	2	2
$P(X=3)=0.037$		0	3	7
	總數	8	18	20



■ Algorithm:

1. Generate  $U \sim U(0,1)$

2. If  $0 \leq U < 0.8$ , let  $I = [10U] + 1$ ,  $X = a_1[I]$

$0.8 \leq U < 0.98$ , let  $I = [100U] - 80 + 1$ ,  $X = a_2[I]$

$0.98 \leq U$ , let  $I = [1000U] - 980 + 1$ ,  $X = a_3[I]$

其中

$$a_1[.] = 00 \ 1111 \ 22$$

$$a_2[.] = 0000000000 \ 1111 \ 22 \ 333$$

$$a_3[.] = 00000 \ 11111 \ 22 \ 33333333$$

- Example 3. Generate an r.v.  $X$  from

$$f(x) = n \int_1^{\infty} y^{-n} e^{-xy} dy, \quad 0 < x < \infty,$$

where  $dF_Y(y) = \frac{n dy}{y^{n+1}}, 1 < y < \infty, n \geq 1.$

- Algorithm:

1. Generate  $U_1, U_2$  from  $U(0,1)$

2. Let  $Y \leftarrow (U_1)^{-1/n}$

3. Return  $X \leftarrow -\log(U_2)/y$

Note:  $dF_Y(y) = \frac{n dy}{y^{n+1}} \Rightarrow F_Y(y) = \frac{1}{y^n} \sim U(0,1).$

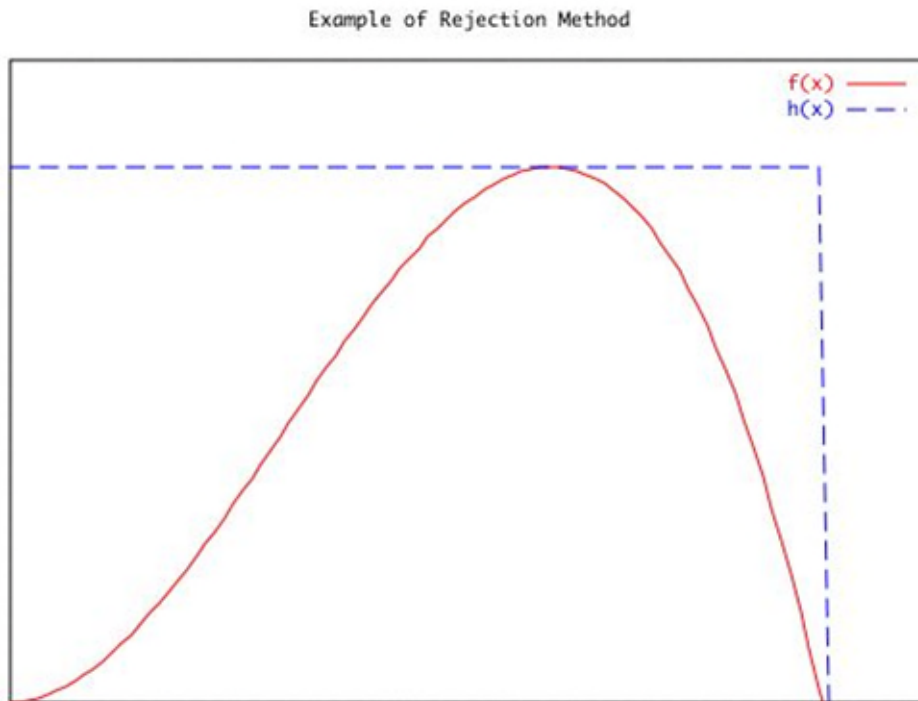


# Rejection Method

Generates random numbers for a distribution function  $f(x)$ . This is achieved as follows:

- Define a comparison function  $h(x)$  such that it encloses the desired function  $f(x)$ .
- Choose uniformly distributed random points under  $h(x)$ .
- If a point lies outside the area under  $f(x)$  reject it and choose another point.

# Illustration of the Rejection Method



The following is an illustration of the rejection method using a square function for the comparison function.

- Example 1.  $X \sim \text{Beta}(2,4)$ , i.e.  $E(X) = 1/3$ ,

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1.$$

Then we use  $g(x) = 1$ ,  $0 < x < 1$  as “envelop” to create  $f(x)$ .

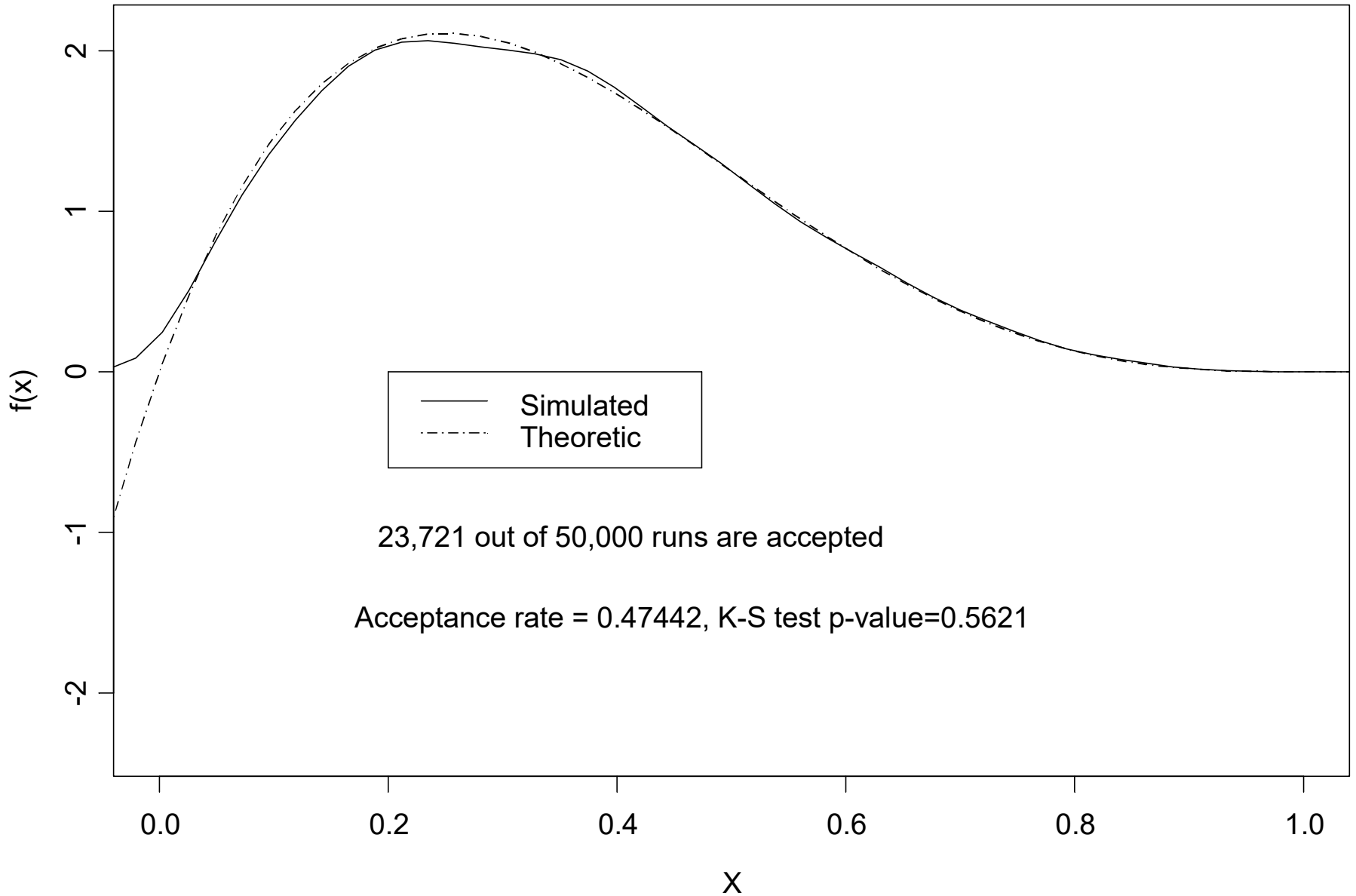
Note:  $\frac{f(x)}{g(x)} = 20x(1-x)^3 \leq \frac{135}{64} = C \approx (0.4741)^{-1}$ .

- Algorithm:

1. Generate  $X$ ,  $U_1 \sim U(0,1)$ .

2. If  $U_1 \leq \frac{f(x)}{C \cdot g(x)}$  return  $Y = X$ . (Q: Rejection rate?)

# Theoretic and simulated $f(x)$ 's for Beta(2,4)



- Example 2. Generate  $\text{Gamma}(3/2, 1)$ , i.e.

$$f(x) = \frac{2}{\sqrt{\pi}} x^{\frac{1}{2}} e^{-x}, x > 0.$$

We want to generate  $X$  from  $g(x) = \frac{2}{3} e^{-2x/3}$ .

The max. of  $f(x)/g(x)$  is obtained when

$$\frac{1}{2} x^{-1/2} e^{-x/3} = \frac{1}{3} x^{1/2} e^{-x/3} \Rightarrow x = \frac{3}{2},$$

since  $\frac{f(x)}{g(x)} = \frac{3}{\sqrt{\pi}} x^{1/2} e^{-x/3}$ .

$$\therefore C = \frac{3}{\sqrt{\pi}} \left(\frac{3}{2}\right)^{1/2} e^{-1/2} = \frac{3\sqrt{3}}{\sqrt{2\pi e}} \cong 1.257317.$$



■ Algorithm:

1. Generate  $U$ ,  $U_1 \sim U(0,1)$

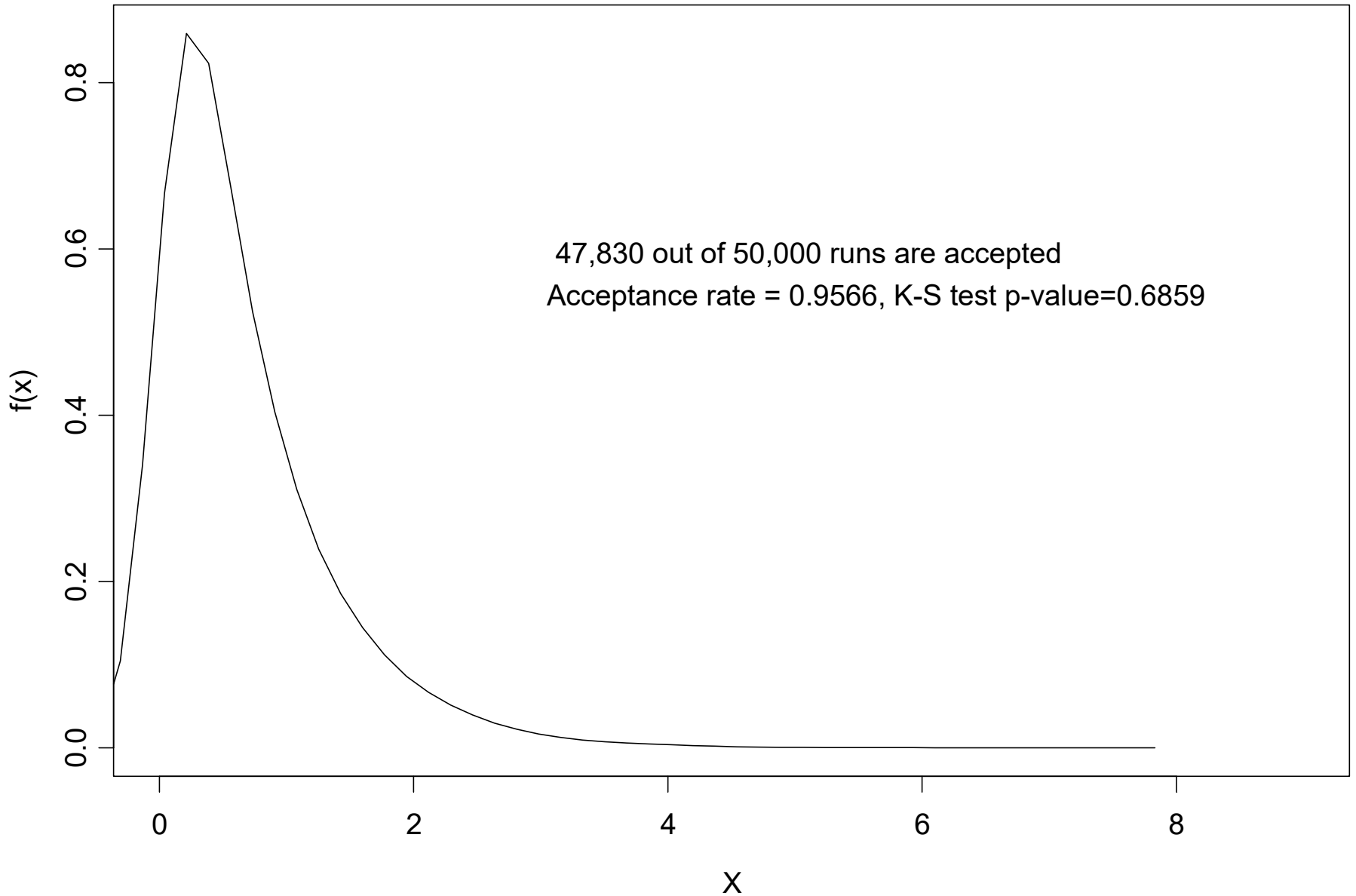
Let  $X = -3 \log U / 2$ .

2. Return  $Y = X$  if  $U_1 \leq \frac{f(y)}{C \cdot g(y)}$ .

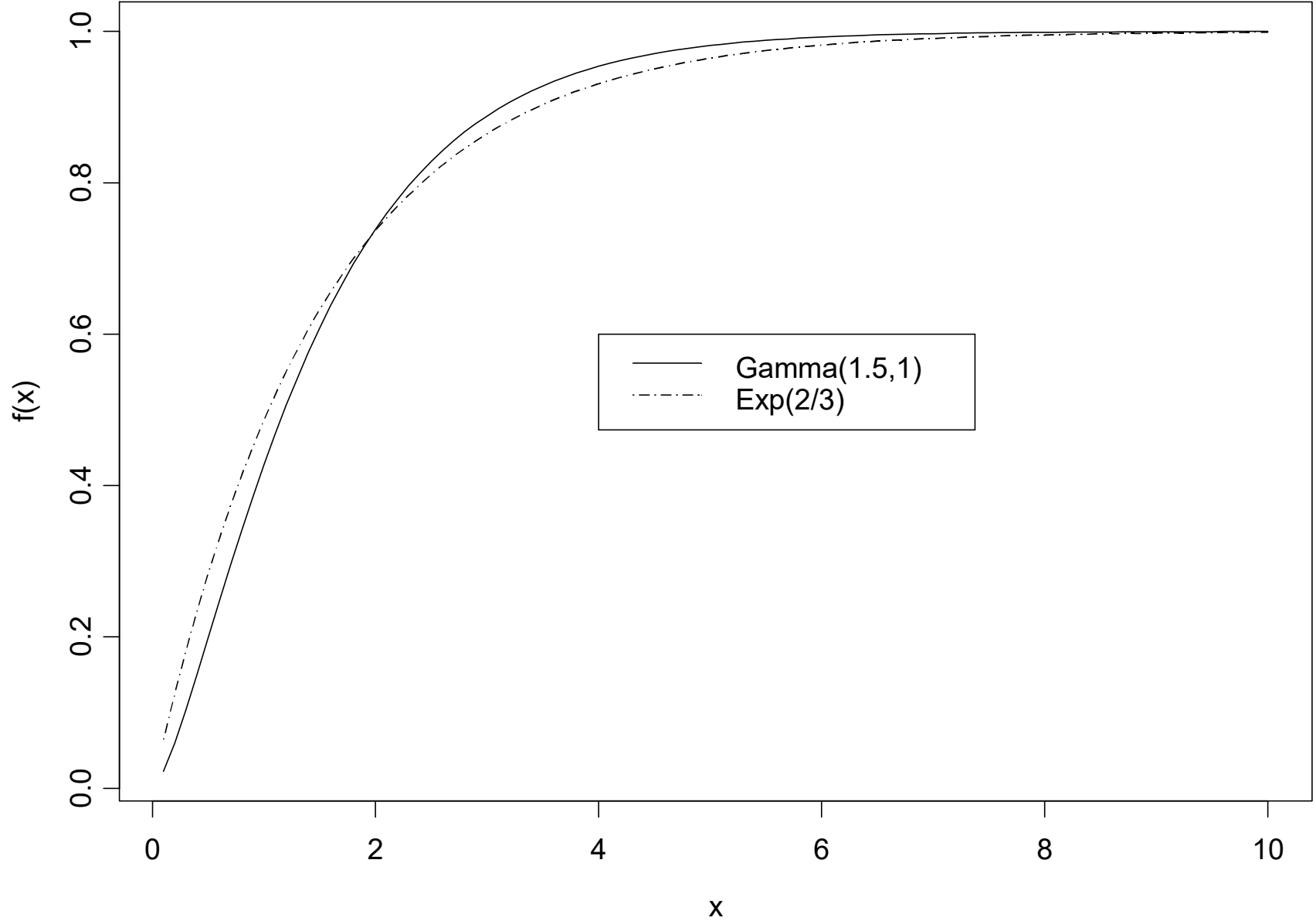
■ Question: Why do we choose  $Y \sim \text{Exp}(2/3)$ ?

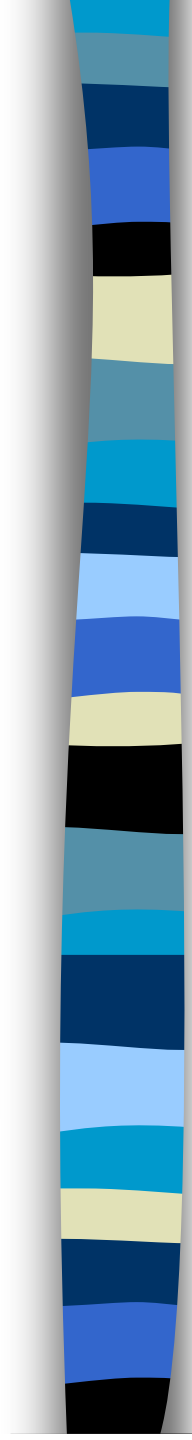
→  $\text{Gamma}(3/2, 1)$  and  $\text{Exp}(2/3)$  have the same mean!

# Simulated $f(x)$ 's for Gamma(3/2, 1)



# Density functions of Gamma(1.5,1) and Exp(2/3)



- 
- Alias method: Looks like “rejection” but it is indeed “composition”.
  - Example 1.  $X \sim B(3, 1/3)$ , i.e.

$$P(X = i) = \frac{8}{27}, \frac{12}{27}, \frac{6}{27}, \frac{1}{27} \text{ for } i = 0, 1, 2, 3.$$

$$P(X = 0) = \frac{1}{4} + \left[ \frac{2/27}{4} + \frac{3/27}{4} \right] = \frac{32}{108};$$

$$P(X = 1) = \frac{25}{108} + \left[ \frac{23/27}{4} \right] = \frac{48}{108};$$

$$P(X = 2) = \frac{24}{108};$$

$$P(X = 3) = \frac{4}{108}.$$

